Structure 14, 511–520, March 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2005.11.020

Cryo-Electron Microscopy Studies of Human TFIID: Conformational Breathing in the Integration of Gene Regulatory Cues

Patricia Grob,[1] Michael J. Cruse,[2] Carla Inouye,[1] Marian Peris,[3] Pawel A. Penczek,[4] Robert Tjian,[1,3] and Eva Nogales[1,2,3,*]

[1] Molecular and Cell Biology Department
University of California, Berkeley
Berkeley, California 94720

[2] Physical Biosciences Division
Lawrence Berkeley National Laboratory
Berkeley, California 94720

[3] Howard Hughes Medical Institute
Molecular and Cell Biology Department
University of California, Berkeley
Berkeley, California 94720

[4] Department of Biochemistry and Molecular Biology
The University of Texas—Houston Medical School
6431 Fannin, MSB 6.218
Houston, Texas 77030

Summary

The multisubunit transcription factor TFIID is essential for directing eukaryotic promoter recognition and mediating interactions with activators/cofactors during assembly of the preinitiation complex. Despite its central role in transcription initiation and regulation, structural knowledge of the TFIID complex has so far been largely limited to electron microscopy studies of negatively stained samples. Here, we present a cryo-electron microscopy 3D reconstruction of the large endogenous human TFIID complex. The improved cryopreservation has allowed for a more detailed definition of the structural elements in the complex and for the detection, by an extensive statistical analysis of the data, of a conformational opening and closing of the cavity central to the TFIID architecture. We propose that these density rearrangements in the structure are a likely reflection of the plasticity of the interactions between TFIID and its many partner proteins.

Introduction

The regulated transcription of eukaryotic protein coding genes by RNA polymerase II (Pol II) requires the coordinated assembly of a set of proteins and multiprotein complexes at specific DNA sequences in the promoter. The transcription factor IID (TFIID), an essential element of this complex machinery, recognizes and binds promoter DNA and is thought to help recruit the other general transcription factors (TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH) and Pol II to form the preinitiation complex (PIC) (reviewed in Orphanides et al., 1996; Roeder, 1996). TFIID is itself a large multiprotein complex that consists of the TATA box binding protein (TBP) and a set of 12–14 TBP-associated factors (TAFs) (Goodrich and Tjian, 1994). TBP alone is able to sustain basal levels of transcription (Kao et al., 1990; Peterson et al., 1990). However, holo-TFIID appears to be required in a number of contexts for activated transcription as well as for binding to promoter sequences other than the TATA box, such as the initiator (Inr) and the downstream promoter element (DPE) (Pugh and Tjian, 1990; Tanese et al., 1991; Dynlacht et al., 1991; Kaufmann et al., 1998; Emami et al., 1997; Martinez et al., 1998; Burke and Kadonaga, 1997; Zhou and Chiang, 2002).

In spite of its instrumental role in transcription initiation, information regarding TFIID's structural organization remains scarce. Conventional X-ray crystallography and NMR studies have revealed the structures of a few individual subunits and subdomains. For example, TBP has been extensively studied by itself or in a complex with promoter DNA and with either TFIIA or TFIIB (Geiger et al., 1996; Tan et al., 1996; Tsai and Sigler, 2000). Some TAFs and TAF pairs, including the so-called histone fold TAFs and several domains of TAF250, have also been characterized structurally (Hoffmann et al., 1996; Selleck et al., 2001; Jacobson et al., 2000; Liu et al., 1998; Mal et al., 2004). However, details on the overall organization of the TBP/TAF complex have been elusive, possibly due to its size and the inherent difficulties in producing large amounts of active assembled complexes. The most useful structural technique able to overcome these difficulties has been electron microscopy and image reconstruction. The first low-resolution, 3D structures of human and yeast TFIID were obtained by single-particle methodology with negatively stained samples (Andel et al., 1999; Brand et al., 1999; Leurent et al., 2002). Although these studies constituted an important step in defining the overall architecture of the complex, the quality of the results was limited by staining and dehydration effects.

Here, we report the first, to our knowledge, structural characterization of frozen-hydrated, native, endogenous human TFIID using cryo-electron microscopy (cryo-EM) and single-particle reconstruction. The improved preservation of the complex allowed us to analyze different classes of TFIID as representing multiple conformational/biochemical forms of endogenous TFIID in solution. While single-particle methodology generally requires (or assumes) that the sample is homogeneous, recent approaches have aimed to analyze and characterize the intrinsic variability present in many macromolecular complexes (see for example Heymann et al., 2004; Gao et al., 2004; Tilley et al., 2005; Zhou et al., 2001; Brink et al., 2004). Here, we have used 2D and 3D classification and statistical methods to dissect the range of structural flexibility exhibited by human TFIID. We find that the complex is able to alternate between at least two different conformations, so that the relative positions of the three characteristic "architectural" lobes of TFIID change in a coordinated fashion. The conformational variability described here may represent functionally distinct TFIID complexes important for regulation of transcriptional activation in metazoan organisms.

*Correspondence: enogales@lbl.gov

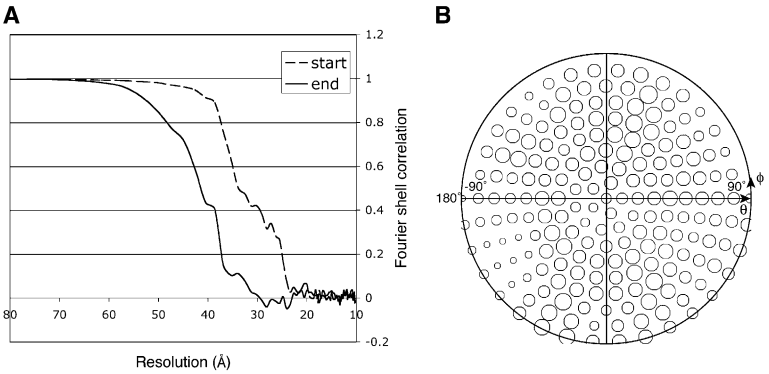

Figure 1. Refinement of the 3D Reconstruction by Projection Matching

(A) Fourier shell correlation curves measuring the resolution at the beginning and the end of the refinement. The 0.5 threshold indicates a final resolution of 32 Å.

(B) Angular distribution plot of the particles selected for the final 3D reconstruction showing a fairly isotropic distribution of orientations (the circles are proportional to the number of particles).

Results

Cryo-EM 3D Reconstruction of Human TFIID

Our structural studies have been carried out by using endogenous human TFIID immunopurified from HeLa cell nuclei that is transcriptionally active. Images of frozen-hydrated complexes were recorded in a microscope equipped with a field emission gun (FEG) using far defocus settings in order to overcome the contrast reduction due to the buffer composition (presence of glycerol and sugars) and the presence of a thin carbon support (required to increase the concentration of complexes on the grid). Several cycles of 2D classification and alignment with IMAGIC (van Heel et al., 1996) were employed in order to progressively sort out a final data set of 12,816 particles into different views of the complex. The first cryo-EM model was obtained from the best class averages by using common lines in IMAGIC with several reprojections of the previous negatively stained model (Andel et al., 1999) as an initial anchor set. After a few cycles of refinement, the resolution stopped improving at 40 Å. The model was then further refined by a modified iterative projection matching procedure using SPIDER (Frank et al., 1996). The final resolution after 12 cycles of refinement was 32 Å, as determined by the Fourier shell correlation (FSC) criterion with a cutoff value of 0.5 (Saxton and Baumeister, 1982) (Figure 1A). The angular distribution of the directions of projection data (see Figure 1B) is fairly isotropic.

The refined model of TFIID in Figure 2 is filtered to 32 Å resolution and is displayed with a threshold chosen to include a mass of 1 MDa. This is an approximate threshold, as the precise mass of human TFIID has not been accurately determined (see Discussion). Overall, the structure has the general "horseshoe" shape observed in our previous reconstruction from negatively stained samples (Andel et al., 1999). However, our new structure has gained several distinct features and an improved spatial appearance. The diameter of the structure is about 200 Å, and it has three main lobes of similar size, A, B, and C, arranged around a large central cavity shaped by a number of surrounding channels. This labeling of the main density domains corresponds to that of the negative stain, as obtained by visual alignment of the structures, since both models were obtained independently in order to avoid model bias. In what we call a "front view" (Figure 2A) the viewer looks down the central cavity, which is open at the bottom and surrounded by lobes A, C, and B, arranged from right to left in an anticlockwise fashion. Lobes A and B connect to C through bridges of density on opposite sides of the complex and at different heights, defining the major side channels into the main cavity (Ch1 and Ch2 in Figures 2A and 2B). Lobe C is more indented than the globular A and B lobes, and it has two spikes of density at the front surface of the complex (c1 and c2), best seen in the "top" view (Figure 2B). Connected more loosely to C are two minor lobes, d1 and d2, on the bottom surface of the complex, most clearly seen in the "bottom" view (Figure 2C). These extensions, which give added depth to the structure and were barely apparent in the previous negatively stained reconstructions, increase considerably the solvent-accessible surface and contribute to the complex shape of the central cavity as well as the channels that connect it to the outside. The main cavity opening is defined by the space between lobes A and B, about 40 Å wide at its narrowest point. Two side channels are noticeable: Ch1, a long, narrow groove in the back of the structure, between lobes B, C, and d2; and Ch2, a smaller, rounder channel between lobes A, C, and c1, which, together with the central cavity, make a continuous water-accessible path all the way across the complex (most clearly seen in Figure 2C). Also visible is a smaller opening, Ch3, at the back of the structure between lobes C, d1, and d2, most clearly seen in the "back" view (Figure 2E).

Structural Variability of Human TFIID

The somehow limited resolution of the structure given the size of our data set, as well as our analysis of 2D class averages, suggested that the complex was present in more than one biochemical or conformational state. To characterize this variability, we subclassified within each of the 84 projection classes containing particles corresponding to a similar view of the complex (15° angular step size). Multivariate statistical analysis (MSA) and classification of each projection group confirmed the structural variability of the complex. In particular, different relative positions of lobes A and B were revealed and were accompanied by the sporadic appearance of additional mass in the center of the complex (see later). In light of these results, we decided to compute a 3D real space variance map by using a new bootstrap method (Penczek et al., 2006). Two variance maps were obtained from 500 reconstructions, each using resampled data sets for particles and for samples of the

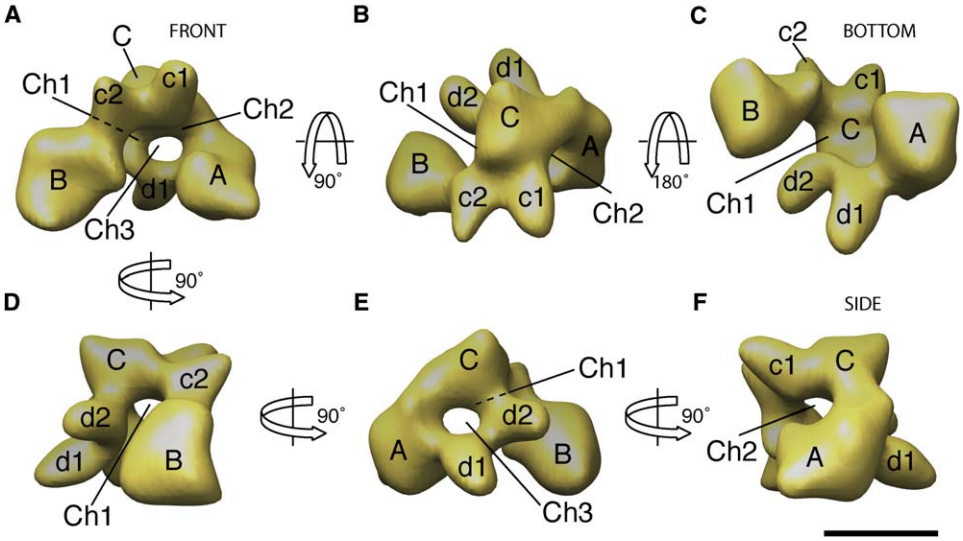

Figure 2. 3D Reconstruction of Frozen-Hydrated Human TFIID at 32 Å

(A–F) The density threshold has been chosen for a 1 MDa complex. The main lobes of density have been labeled A, B, and C, and the smaller ones are labeled c1, c2, d1, and d2. The channels and cavities have been labeled Ch1, Ch2, and Ch3. (A) Front view. (B) Top view. (C) Bottom view. (E) Back view. (D and F) Side views. The scale bar represents 100 Å.

background noise, respectively. The variance values for the structure were significantly higher than the background variance in some regions of the structure, indicating the presence of significant variation in our data set (Figure 3). These regions spanned part of the central cavity and the two main side channels. Additional "hot spots" were present in the center of lobes A and B. In order to classify the data set according to the variance hot spots, a 3D mask was generated enclosing the high-variance region (red in Figure 3) that was then projected in 84 quasi-evenly distributed angular directions. Each projection group was classified by MSA within its respective mask. Although initially we generated up to five subclasses for each projection direction, visual inspection indicated that the majority of the data can be best described by sorting the particles into just two classes, corresponding to two main conformations, one "full" or "closed" and one "empty" or "open" (Figure 4E). The data set was consequently separated into two groups, from which two 3D reconstructions were computed and then refined by multireference projection matching (see Experimental Procedures). The particles were free to align to either model during the refinement, and 57% of them correlated best to volume 1 ("closed"), while 43% did so to volume 2 ("open"). The open conformation (volume 2) resembles the initial, average model, but with a wider, more open central cavity (Figure 5B). Lobes A and B are farther apart, and the rotation of lobe A in the opposite direction to that of d2 leads to the opening of the main entrance to the central cavity (~50 Å; best seen in the bottom view of Figure 5B). Channel 1 between lobes A and c1 also opens as a result of the A rotation, and it is accompanied by a movement of c1 in the opposite direction (best seen in the front view of Figure 5B). In the closed conformation (volume 1, Figure 5A), the concomitant closure of this channel leads to a new contact between the front end of A and lobe c1 (Figure 5A, front view). As a result, the central cavity of the closed conformation is more "pinched": d2 reaches inside the cavity as lobe A is rotated towards it. The difference map between the closed and open conformations (Figure 5C) shows that the red densities corresponding to "positive" differences almost always have a corresponding green, "negative" difference, supporting the general idea of a conformational change between volumes 1 and 2 ("open/closed"), rather than a major difference in biochemical composition (the "full/empty" alternative). Most of the 3D variance map can be superimposed to the features in the difference map, further supporting the hypothesis that the 3D variance originates primarily from small movements of the different lobes. However, small biochemical differences in composition cannot be entirely ruled out. For example, there is no clear matching red density for the green one on the outer surface of lobe B, which seems to have a different shape in both conformations. It is also unclear whether some of the "red" density difference in the center of the complex could be due to the presence of small additional mass in that area. We calculated the FSC curves for each of the two groups, both directly or adjusted to the size of the entire data set (see Experimental Procedures). The resolution determined by the threshold at 0.5 of the FSC curve is 32 Å for the entire data set, and it is 33 Å and 35 Å for group 1 and group 2, respectively, before adjustment. The adjusted resolutions are 30 Å for group 1 and 32 Å for group 2. It appears that the separation of the databases on the density within the high-variance central cavity of the complex resulted in a partition into what we interpret to be a slightly more stable conformation (1) and a slightly more flexible one (2). The 3D variance calculated after partition also seems to confirm this observation. The average variance in reconstruction 1 is lower than in a random group of the same size picked from the entire set, while the variance in reconstruction 2 is somewhat higher (data not shown). This effect is likely due to the fact that the closed

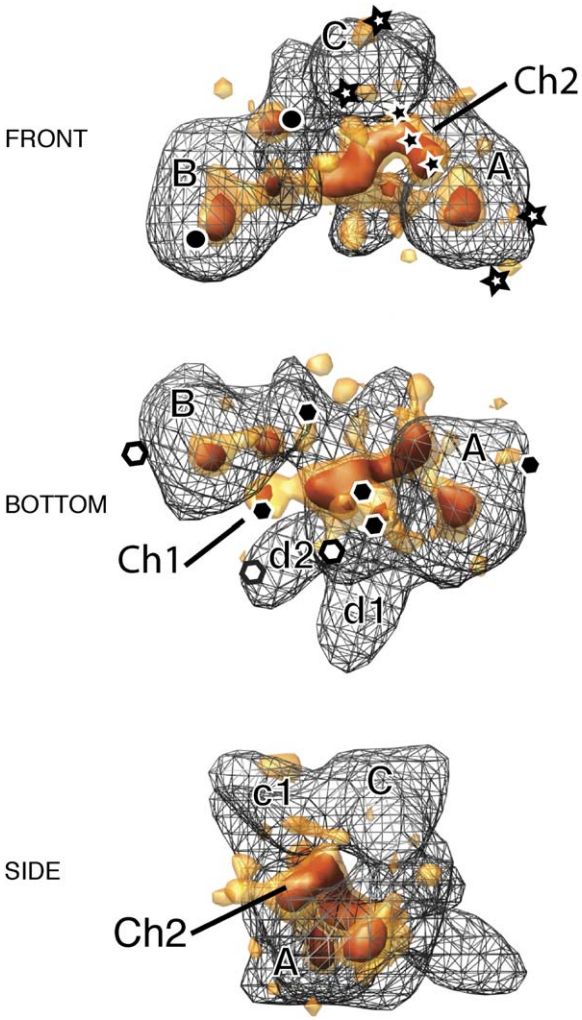

Figure 3. 3D Variance Map and Covariance Analysis

The high density (red) and lower density (yellow) of the 3D variance map are superposed onto the 3D reconstruction shown in Figure 2 (mesh). The high variance is mainly localized inside the central cavity, Ch1, and Ch2. A few hot spots are also visible at the tips of lobes A and B, around their hinges, and around c1, c2, d1, and d2. The results of the covariance analysis are represented by the black and white symbols. Each symbol (star, hexagon, and dot) corresponds to a group of variance points that have a significantly high correlation between each other (see Experimental Procedures). Black-white represents a negative correlation, and white-white and black-black represent a positive correlation.

conformation appears to have additional protein-protein contacts compared to the open conformation, thus stabilizing the closed structure. This result also supports the idea of conformational breathing, with the present group 2 still including a range of conformations. A much larger data set could therefore allow us to isolate more conformational intermediates in the future.

Covariance Analysis

To investigate if the movements described above are indeed correlated, we analyzed the covariance between high-variance regions of the density map. There was significant positive and negative correlation between densities localized in different lobes of the 3D structures, indicating concerted movements, much in the same way as the positive and negative parts of the difference map (symbols on the 3D variance map Figure 3). Some of these movements appeared to cluster in three different groups: (i) the rearrangement of density at the A-d1 hinge seems to be associated with an inward displacement of the d1-d2 density, and with an adjustment of lobe B about its hinge with C (likely mediated by contacts between d2 and B) (hexagons, Figure 3, bottom view); (ii) the closure of channel 2 is coupled with both a downward movement of c1 and an upward and forward movement of A (stars); (iii) variations in the hinge region between B and C are linked to the position/conformation of the tip of lobe B (dots). It was not clear if groups (i) and (ii) were correlated through the movement of lobe A. These results confirm the existence of concerted movements in the complex.

Discussion

New Structural Features of Human TFIID Revealed by Cryo-EM Studies

Due to difficulties in obtaining sufficient quantities of native, active, human TFIID, as well as other multisubunit complexes involved in eukaryotic transcription initiation, the only 3D structures of these essential gene regulatory protein complexes have been 3D EM reconstructions with negatively stained samples. This sample preparation method is simpler, requires small amounts of material, can be more permissive to certain buffer conditions, and yields high-contrast images that facilitate the alignment of the data. However, negative staining can induce artifacts due to flattening during dehydration and imperfect staining that result in an unfaithful representation of the 3D organization of the macromolecular complex under study. By contrast, the frozen-hydrated preparation of protein complexes (cryo-EM) preserves the structure of proteins in a close-to-native state and thus allows for the detection of conformational variability that may be important for the biological function of the complex under examination.

Our present study of frozen-hydrated human TFIID reveals features that generally agree with our previous negatively stained reconstruction, namely, the subdivision of the complex into three large lobes arranged around a central cavity. Visual comparison of our TFIID cryo-EM structure with the previous model from negative stain allows us to approximately locate TBP and the binding sites for TFIIA and TFIIB. TFIIB would be binding in the neighborhood of channel 1 and TFIIA between the back of lobe A and d1, while TBP would be located by the main cavity, in contact with lobe C. On the other hand, new density features are visible in the cryo-EM reconstruction, represented by the c1 and c2 protrusions off lobe C and the smaller lobes d1 and d2, which were barely hinted at in the previous study (Andel et al., 1999). There are also significant differences in the location/definition of the contacts between the lobes. The connections between density C and lobes A and B are both located toward the back of the negatively stained reconstruction, whereas they alternate—back for A-C and front for B-C—in the cryo-EM model. These differences are likely to be caused by the penetration and coverage of the stain in our previous structure, as well

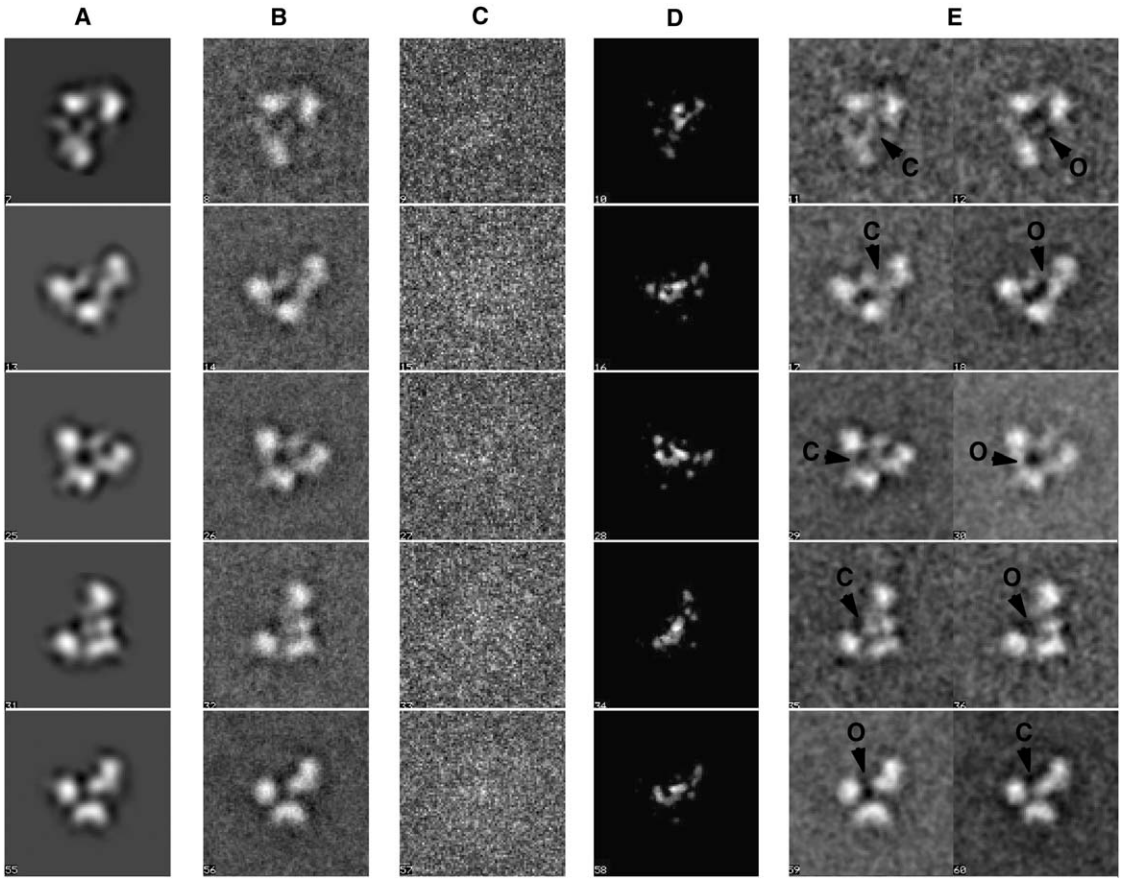

Figure 4. 2D Classification within the High-Variance Masks

(A–E) Each row corresponds to one of a few representative views of the complex. (A) Reprojections of the 3D reconstruction. (B) Total 2D average of the particles in the same view. (C) Corresponding 2D variance. (D) Reprojections of the 3D variance map. (E) New class averages. The arrows point out the differences between the "closed/full" classes (c) and the "open/empty" ones (o) that were initially used to separate the particles into two groups.

as by the effect of partial flattening for certain orientations on the support surface.

The overall architecture of the better-preserved, frozen-hydrated complex resembles a short spiral enveloping the empty center (see bottom and side views in Figure 2) rather than a flat horseshoe, with the connections between A and B and the central C lobe likely playing the role of hinges. The large central cavity can easily accommodate promoter DNA and, most significantly, a nucleosome, particularly in the open conformation (see Figure 5B, bottom view). The newly defined channels in the structure could also allow a selected number of substrates or target molecules to enter and exit the complex at different stages of transcription. The extensive and corrugated surface of the TFIID complex is consistent with the variety of factors that are thought to interact with TFIID during transcriptional activation and PIC formation.

Structural Breathing of Endogenous Human TFIID

Because protein complexes are studied free in solution, single-particle image reconstruction methodology is "limited" by the potential conformational flexibility of the complex, which, of course, is obviated in X-ray crystallographic studies. However, if the appropriate methodology is utilized, single-particle reconstruction can detect and even characterize variability in the sample that may lead to a more representative structural description of large flexible complexes such as TFIID. Due to the preponderance of X-ray crystallographic studies, the potential significance of conformational flexibility in macromolecular assemblies remains poorly understood. However, the possibility and magnitude of such flexibility in large complexes will not come as a surprise, particularly for those whose functions involve multiple and transitory interactions with many target factors, as is the case for eukaryotic transcriptional regulatory complexes. Furthermore, for these large complexes that have the capacity for amplifying biochemical reorganization at the atomic level into large movements, the resolution attainable by cryo-EM reconstruction can be particularly well suited. As an example within the transcriptional field, the coactivator CRSP/Mediator can dramatically change conformation, in an activator-dependent manner, at a scale that makes it possible to be characterized by relatively low-resolution methods such as EM (Taatjes et al., 2002, 2004). In the present study, a unique biochemical preparation gave rise to at least two distinct conformations of the large human TFIID complex. Discriminating this conformational

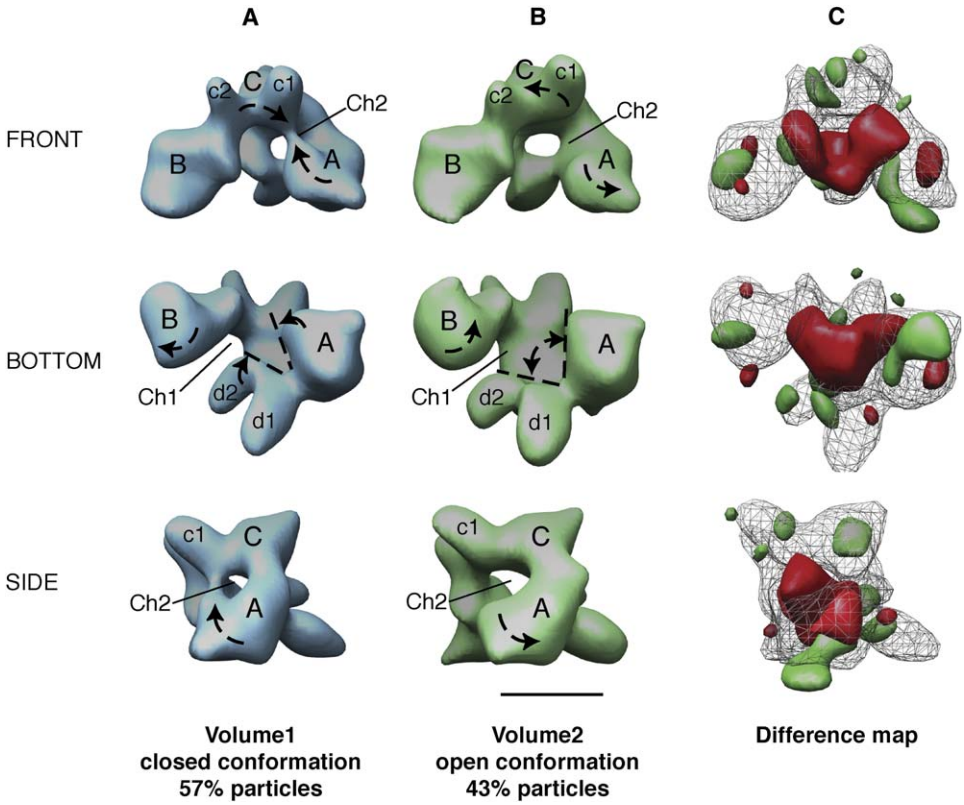

Figure 5. Alternative Conformations of Endogenous Human TFIID

(A–B) Two distinct 3D reconstructions were obtained from the two particle groups separated by classification within the high-variance region. (A) Closed conformation obtained from the first group of 2D class averages (Figure 4E). (B) Open conformation from the second group of class averages (Figure 4E). Both models are represented with a density threshold corresponding to a 1MDa protein complex, in the same orientation on each row (superposed in Chimera, see Experimental Procedures). The lines and arrows point out the different "movements" described in the text. (C) Difference map between the two conformations. The positive part (red) and negative part (green) of the (Volume1 − Volume2) difference are superposed on the average of the two volumes (mesh). The average volume is identical to the original reconstruction obtained from the entire data set. The scale bar represents 100 Å.

variability was made possible by extensive data processing, particularly calculation of variance maps in 2D and 3D and application of a dedicated classification procedure.

Single-particle reconstruction relies on averaging of massive amount of data, both in 2D and in 3D. Thus, in addition to 2D averages and the 3D density map, important information is contained in the variance. The analysis of variance can provide clues concerning the homogeneity of the data and the structural significance of the observed fluctuations. Classification of images and variance analysis in 2D has traditionally been used to detect and interpret interimage variability in electron microscopy. In some cases, it was even possible to reconstruct different conformational states present in a mixed population in 3D from the information available in 2D, but the methods used were not general and were usually specific to the particular case studied (Heymann et al., 2004; Gao et al., 2004; Tilley et al., 2005; Zhou et al., 2001; Brink et al., 2004). In this study, we used the bootstrap technique (Penczek et al., 2006) in order to estimate directly the real space variance in 3D to characterize the conformational variability of human TFIID. The 3D covariance has added a new dimension to the analysis, allowing us to identify coordinated movements within the TFIID complex.

A number of possibilities could explain the presence of more than one conformational state in our HeLa cell-purified TFIID samples. Different forms of TFIID have been reported to coexist in HeLa cells, with different transcriptional properties. Complexes with alternative TAF subunit compositions have been isolated. One of these subforms is the TAF30-containing TFIID, which lacks TAF18 and is thought to play a role in estrogen receptor activation (Jacq et al., 1994). More recently, a subpopulation of TFIID, containing the B cell-specific factor TAF105, has been identified that may also be present at very low levels in other human cell lines (Dikstein et al., 1996a). The TAF stoichiometry has not been clearly established for human TFIID, and it may be another source of variability (R. Coleman, personal communication) (Selleck et al., 2001; Sanders et al., 2002). Additionally, different subunits of TFIID are direct targets for gene-specific activators and inhibitors of transcription (Pugh and Tjian, 1990; Tanese et al., 1991; Dynlacht et al., 1991). Some of these interactions may induce a rearrangement of the complex. For example, in the case of the human coactivator complex CRSP, direct reorganization of the complex upon binding of different activators has been observed and proposed as a mechanism for transcription activation (Taatjes et al., 2002, 2004). It is possible that activated

or inhibited forms of TFIID coexist with a "neutral" conformation, either because the modification is permanent, or because TFIID naturally fluctuates between different states in the absence of ligand. Finally, several posttranslational modifications of TAFs have been identified (Mizzen et al., 1996; Dikstein et al., 1996b; Segil et al., 1996; Kouskouti et al., 2004; Sawa et al., 2004; Boyer-Guittaut et al., 2005). The mechanism and the extent of these modifications in human cells remain unclear, but they could induce various rearrangements of TFIID, as they seem to participate in the dynamic regulation of TFIID promoter and chromatin binding activities.

Cooperative Movement of Structural Domains in the Integration of Transcriptional Cues

The variability that we observe in endogenous human TFIID could have different physical origins. As mentioned above, TFIID appears to exist in different biochemical flavors, not just in different tissues, but also within a single cell type. This compositional variability can involve the swapping of TAFs or a variety of posttranslational modifications, both of which are likely to result in additional conformational variability. Irrespective of whether the conformational flexibility that we see in our study originates from distinct biochemical states or not, our findings suggest a "breathing" mechanism that may be inherent to the architectural organization of the complex into discreet lobes with hinge-like connections. We propose that the 3D variability that we observe in human TFIID could reflect the potential of the complex to reorganize itself for binding to different ligands. A nontrivial finding of our analysis is that the regions of flexibility distributed throughout the density of the complex are coupled, suggesting the potential cooperativity of binding sites. The movements illustrated in our two alternative conformations reveal that the central cavity and channels between lobes can rearrange to change their size and likely their chemical character, giving the complex a versatility of potential binding sites and exposed surfaces that can be cooperatively linked through the scaffold of the structure. This versatility may reflect an intrinsic property of the complex that could play a distinct role in directing the formation of an active PIC at different promoters through interactions of TFIID with different sets of activators and cofactors.

Experimental Procedures

Cryo-Electron Microscopy of Human TFIID

The human TFIID complex was immunopurified from HeLa cell nuclear extracts as previously described (Naar et al., 1998). Both initial electron microscopy imaging of negatively stained samples and SDS-PAGE analysis indicated the presence of a variable subunit composition in the preparations that was especially apparent when samples were repeatedly freeze-thawed. Consequently, all of the data used for this study were collected from a single, fresh preparation of TFIID. 400 mesh copper grids were covered with a holey carbon film (Bradley, 1965) and by a thin layer of continuous carbon. A fresh TFIID preparation in 10% glycerol was diluted down to about 25 nM in HEMK buffer (25 mM HEPES [pH 7.9], 0.1 mM EDTA, 12.5 mM $MgCl_2$, 200 mM KCl, 0.03% NP-40) containing 3% trehalose for complex stabilization. A freshly glow-discharged grid was mounted in a Vitrobot chamber set at 95% relative humidity and 10ºC. A total of 4 μl of the protein sample was applied to the grid for 30 s, then rinsed in 4 μl HEMK to remove the excess trehalose before vitrification in liquid ethane (Adrian et al., 1984). Cryo-electron microscopy images were recorded on a Philips CM200 microscope equipped with a field emission gun, operating at an acceleration voltage of 200 kV and at a magnification of 50,200×. Data were collected in low-dose conditions (approximately 16 electrons/Å^2) on Kodak SO163 films with an underfocus ranging from 3.0 to 5.5 μm.

Image Preprocessing

Micrographs were digitized in a Nikon Super Coolscan 8000 scanner at a pixel size of 12.7 μm, resulting in a pixel size of 2.53 Å on the specimen scale, subsequently decimated to yield 5.06 Å/pixel. For the purpose of the particle picking, in order to partially overcome the low contrast due to both the sample buffer and the presence of the thin carbon support, appropriate high-pass and low-pass filtering was applied to micrographs. Particles were selected interactively from the filtered images using Boxer (EMAN [Ludtke et al., 1999]), then extracted from the unfiltered images using a window size of 120 × 120 pixels. The final data set after visual culling reached a total of 12,816 particles. A large enough window size was chosen to facilitate the determination of the contrast transfer function (CTF) parameters by using the Fitctf utility of EMAN and the subsequent CTF correction. The particles were then transferred to SPIDER for normalization of densities, CTF phase correction, low-pass filtering, and extraction of smaller 80 × 80 pixel particle boxes. The data set was then converted to IMAGIC format for initial analysis.

Initial 3D Reconstruction in IMAGIC

Images were centered and subjected to five successive rounds of multivariate statistical analysis (MSA), and multireference alignment (MRA) in IMAGIC (van Heel et al., 1996; Stark et al., 1995). Classification and alignment gave rise to a very large number of distinct classes. The resulting class averages were compared to projections of the negative stain model of TFIID (Andel et al., 1999) after filtering it to 45 Å resolution. The class averages that agreed best with the preexisting model, as determined by crosscorrelation, and were the most homogenous according to visual inspection, were selected as references for the next cycle of MRA. This procedure only aimed at discriminating between "views" of holo-TFIID and classes arising from heterogeneity in the data, and the alignment parameters with the negative stain model were ignored. Additionally, the worst 15%–20% of the data in each class was rejected throughout the process to further reduce the variability within the "good" classes. After 5 iterations, the best 183 class averages were selected to generate a first 3D reconstruction using common lines and back projection with a small random set of projections of the negative stain structure as the initial anchor set. The angular assignment of the 2D averages was refined against reprojections of this new cryo-EM 3D model. A new set of MRA references was generated by reprojecting the refined model every 20º to create a new anchor set. After MRA and MSA, a new angular assignment was obtained, and the process was repeated for 15º, then 10º step size, as the number of class averages participating in the 3D reconstruction was progressively increased (Stark et al., 1995). The resolution of the model obtained was limited to 40 Å resolution as determined by the Fourier shell correlation (FSC) at 0.5 cutoff level (Saxton and Baumeister, 1982).

Refinement of the 3D Reconstruction in SPIDER

Further refinement of the model was carried out using a classic iterative projection matching scheme in SPIDER (Frank et al., 1992, 1996; Penczek et al., 1994) to which an additional 2D alignment step was added.

The initial IMAGIC volume was filtered to 40 Å resolution, reprojected with an angular step of 20º, and used as a reference for the first cycle of projection matching in SPIDER against the unfiltered, phase-corrected particle images. A new volume was then computed using the new angular assignments. The process was repeated several times. Throughout the refinement process, particles were rejected if their crosscorrelation coefficient with the reference was below a certain threshold, which was determined by visual inspection and control of the FSC curves. The angular step size was decreased from 20º to a final of 5º. Because of the low contrast of the data, it was difficult to align the data unequivocally in the vicinity of angular directions along which the reprojections of the structure were nearly mirror-symmetric. Therefore, an additional alignment step was

added between successive steps of 3D projection matching in order to correct the centering of particle projections and to improve the signal-to-noise ratio of the data by averaging. In this step, 2D alignment and clustering of the particle projections assigned to the same angular direction were performed by using the rotationally invariance K-means algorithm (command AP CM in SPIDER) (Penczek et al., 1996). Subsequently, new Eulerian angles were assigned to the class averages, and a corrected reference structure was computed by using the iterative reconstruction algorithm. The final reconstruction included data from 8,016 particles (out of the initial set of 12,816), and its resolution was 32 Å, as evaluated by the FSC criterion (Saxton and Baumeister, 1982; Harauz and van Heel, 1986) at the 0.5 cutoff level (Figure 1A).

2D and 3D Variance Analysis

In order to characterize the heterogeneity of the data, particles were grouped according to their 3D orientation resulting from projection matching into "projection classes" encompassing 15° of the Euler sphere. Every one of these projection groups was subclassified using hierarchal ascendant classification (HAC) within different masks in SPIDER. A first mask encompassing the entire complex indicated variations in three main areas: lobes A and C, and the area in between. A more detailed analysis of the variation in the data set specifically within these areas was pursued by creating a smaller, spherical mask around each one of them and projecting it in the orientations of the respective projection groups. HAC of every group was performed inside the corresponding projection masks (not shown).

In an independent approach, a 3D variance map of the 3D reconstruction was calculated using the technique described by P.A.P. et al., using the program VA 3R (SPIDER) (Penczek et al., 2006).

Briefly, 500 bootstrap versions of the data set, with phases already corrected for the effects of the CTF, were picked randomly. A 3D reconstruction of each of these data sets was computed using the predetermined Euler angles from the iterative reference projection matching. The bootstrap estimate of the 3D variance $\sigma^2_{B_s}$ was calculated using B = 500 resampled volumes filtered at 37 Å. This low-pass filter was chosen to be slightly above the resolution limit of the reconstruction and to reduce the variance that may come from the defocus variations in the data. The same procedure was used to calculate the bootstrap variance of the background noise, from "noise particles" extracted from the area surrounding the particles forming the data set. The real space 3D variance was then calculated from the bootstrap variances and the number of particles, *K*, using the program VA 3R (SPIDER):

$$\sigma^2_{Struct} = K(\sigma^2_B - \bar{\sigma}^2_{Back}), \tag{1}$$

where $\bar{\sigma}^2_{Back}$ is the average of the background variance estimated by the bootstrap technique.

The resulting map is presented in Figure 3, superposed onto the TFIID structure. A 3D mask was generated from the high values of the variance map and was projected every 15° in the same directions as the projection groups. The data inside this mask were subjected to MSA; only the four highest order eigenvectors were used, resulting in a new classification of the data. Visual inspection of the class averages and the repartition of particles between classes revealed that two classes were sufficient to describe most of the variability in the data set. In every view, the two class averages would differ in a similar way, one apparently having higher density in the center cavity of the complex than the other (Figure 4). The data set was accordingly split into two groups by measuring the central density in the class average, and two 3D reconstructions were calculated. The two models obtained from the separation of the data were simultaneously refined by iterative projection matching with the entire data set; the highest crosscorrelation coefficient determined which group each particle belonged to. Three cycles of refinement were performed, finishing at 10° angular step size, at which point the resolution stopped improving at 32 Å and 35 Å resolution for the group with more and less density in the center, respectively. The first reconstruction included 57% of the particles (volume1), and the second contained 43% (volume2). As an additional test for possible model bias, iterative projection matching of the data in group 1 was performed with volume 2 as the initial reference and vice versa. Conformation 1 was obtained from group 1, and conformation 2 was obtained from group 2, confirming that the different conformations were not an artifact imposed on the data, but reflected real conformational variability (data not shown).

In order to compare the resolution obtained before and after separation of the data into smaller subgroups, we calculated adjusted FSC curves. The spectral signal-to-noise ratio (SSNR) is related to the FSC by the relation (Penczek, 2002):

$$SSNR = 2\frac{FSC}{1 - FSC}, \tag{2}$$

when the FSC is calculated from the data set split in half. Additionally, the SSNR of an average or a 3D reconstruction obtained from *N* images is proportional to *N* (Unser et al., 1987). The FSC from a number of particles, N_i, can be adjusted to what it would be for *N* particles by using the correction factor N/N_i:

$$FSC_{i,adj.} = \frac{FSC_i}{FSC_i(1 - N_i/N) + N_i/N}, \; i = 1, 2. \tag{3}$$

We calculated the FSC curves for the entire data set, group 1, group 2, and group 1 and group 2 adjusted to the size of the entire data set. In each case, the resolution was determined by the threshold at 0.5 of the FSC curve. In order to evaluate the effect of the data partition on the structure variance, 3D variance was also calculated for groups 1 and 2 and was compared to the 3D variance from random groups of equal sizes, chosen from the entire data set.

Covariance

The covariance between voxels of the density map can also be calculated with the *B* bootstrap data sets, by using the program VA 3C (SPIDER). This program calculates the crosscorrelation coefficients for a given voxel of interest, *i*:

$$r_{ij} = \frac{\frac{1}{B}\sum_{l=1}^{B}\left(v^l_i - \bar{v}_i\right)\left(v^l_j - \bar{v}_j\right)}{\sigma_{B_i}\sigma_{B_j}}, \tag{4}$$

where *i* and *j* are two voxel indices in a bootstrap 3D map, and v^l_i and $\bar{v}_i$ are the value of the i^{th} voxel of the l^{th} density map and of its average, respectively; the program then stores these values as a 3D map.

The highest value voxels in the 3D variance map were selected, and their covariance maps were calculated. The highest covariance values, between 0.13 and 0.25, indicated significant positive and negative correlation between several high-variance areas, shown by symbols on the 3D variance map (Figure 3).

Difference Maps

Volumes 1 and 2 were aligned as a result of the two-reference projection matching refinement described in the section above. Their 3D densities were normalized and subtracted from one another to obtain two difference maps. Volume1 − Volume2 is represented in red in Figure 5C; Volume2 − Volume1 is represented in green by using isosurfaces corresponding to the same density level . The average of Volume1 and Volume2 was superposed to both difference maps for localization of the differences.

Volume Rendering

All the volumes shown were filtered at the resolution determined by the FSC criterion at the 0.5 cutoff using a Butterworth filter in SPIDER (cutoff ± 0.05 reciprocal unit). The TFIID reconstructions were displayed using a density threshold corresponding to a protein mass of 1MDa. All 3D volumes were rendered with the UCSF Chimera package (Pettersen et al., 2004) from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

Acknowledgments

We thank Robert Coleman and Sacha de Carlo for comments on the manuscript, Agustin Avila-Sakar for electron microscopy support, Slaton Lipscomb for computational support, and Jie Fan for HeLa cell culture. This work was supported by grants from the National Institutes of Health (NIH) General Medical Sciences (P.A.P and E.N.),

the NIH Cancer Research Institute (R.T.), and by the U.S. Department of Energy in contract with the University of California (E.N.). R.T. and E.N. are Howard Hughes Medical Institute Investigators.

Received: September 14, 2005
Revised: November 11, 2005
Accepted: November 16, 2005
Published online: March 14, 2006

References

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32–36.

Andel, F., 3rd, Ladurner, A.G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-dimensional structure of the human TFIID-IIA-IIB complex. Science *286*, 2153–2156.

Boyer-Guittaut, M., Birsoy, K., Potel, C., Elliott, G., Jaffray, E., Desterro, J.M., Hay, R.T., and Oelgeschlager, T. (2005). SUMO-1 modification of human transcription factor (TF) IID complex subunits: inhibition of TFIID promoter-binding activity through SUMO-1 modification of hsTAF5. J. Biol. Chem. *280*, 9937–9945.

Bradley, D.E. (1965). The preparation of specimen support films. In Techniques for Electron Microscopy, D. Kay, ed. (Oxford, UK: Blackwell Scientific Publications), pp. 58–74.

Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science *286*, 2151–2153.

Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., and Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. Structure *12*, 185–191.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. Genes Dev. *11*, 3020–3031.

Dikstein, R., Zhou, S., and Tjian, R. (1996a). Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130. Cell *87*, 137–146.

Dikstein, R., Ruppert, S., and Tjian, R. (1996b). TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. Cell *84*, 781–790.

Dynlacht, B.D., Hoey, T., and Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell *66*, 563–576.

Emami, K.H., Jain, A., and Smale, S.T. (1997). Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. Genes Dev. *11*, 3007–3019.

Frank, J., Penczek, P., and Liu, W. (1992). Alignment, classification, and three-dimensional reconstruction of single particles embedded in ice. Scanning Microsc. Suppl. *6*, 11–20.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. *116*, 190–199.

Gao, H., Valle, M., Ehrenberg, M., and Frank, J. (2004). Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. J. Struct. Biol. *147*, 283–290.

Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast. TFIIA/TBP/DNA complex. Science *272*, 830–836.

Goodrich, J.A., and Tjian, R. (1994). TBP-TAF complexes: selectivity factors for eukaryotic transcription. Curr. Opin. Cell Biol. *6*, 403–409.

Harauz, G., and van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. Optik *73*, 146–156.

Heymann, J.B., Conway, J.F., and Steven, A.C. (2004). Molecular dynamics of protein complexes from four-dimensional cryo-electron microscopy. J. Struct. Biol. *147*, 291–301.

Hoffmann, A., Chiang, C.M., Oelgeschlager, T., Xie, X., Burley, S.K., Nakastani, Y., and Roeder, R.G. (1996). A histone octamer-like structure within TFIID. Nature *380*, 356–359.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422–1425.

Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. Cell *79*, 107–117.

Kao, C.C., Lieberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R., and Berk, A.J. (1990). Cloning of a transcriptionally active human TATA binding factor. Science *248*, 1646–1650.

Kaufmann, J., Ahrens, K., Koop, R., Smale, S.T., and Muller, R. (1998). CIF150, a human cofactor for transcription factor IID-dependent initiator function. Mol. Cell. Biol. *18*, 233–239.

Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol. Cell *14*, 175–182.

Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., and Schultz, P. (2002). Mapping histone fold TAFs within yeast TFIID. EMBO J. *21*, 3424–3433.

Liu, D., Ishima, R., Tong, K.I., Bagby, S., Kokubo, T., Muhandiram, D.R., Kay, L.E., Nakatani, Y., and Ikura, M. (1998). Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. Cell *94*, 573–583.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Mal, T.K., Masutomi, Y., Zhang, Z., Nakata, Y., Ohta, H., Nakatani, Y., Kokubo, T., and Ikura, M. (2004). Structural and functional characterization on the interaction of yeast TFIID subunit TAF1 with TATA-binding protein. J. Mol. Biol. *339*, 681–693.

Martinez, E., Ge, H., Tao, Y., Yuan, C.X., Palhan, V., and Roeder, R.G. (1998). Novel cofactors and TFIIA mediate functional core promoter selectivity by the human TAFII150-containing TFIID complex. Mol. Cell. Biol. *18*, 6571–6583.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261–1270.

Naar, A.M., Beaurang, P.A., Robinson, K.M., Oliner, J.D., Avizonis, D., Scheek, S., Zwicker, J., Kadonaga, J.T., and Tjian, R. (1998). Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. Genes Dev. *12*, 3020–3031.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev. *10*, 2657–2683.

Penczek, P.A. (2002). Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. J. Struct. Biol. *138*, 34–46.

Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy *53*, 251–270.

Penczek, P.A., Zhu, J., and Frank, J. (1996). A common-lines based method for determining orientations for N > 3 particle projections simultaneously. Ultramicroscopy *63*, 205–218.

Penczek, P.A., Yang, C., Frank, J., and Spahn, C.M.T. (2006). Estimation of variance in single-particle reconstruction using the bootstrap technique. J. Struct. Biol., in press. Published online February 13, 2006. 10.1016/j.jsb.2006.01.003.

Peterson, M.G., Tanese, N., Pugh, B.F., and Tjian, R. (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. Science *248*, 1625–1630.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pugh, B.F., and Tjian, R. (1990). Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell *61*, 1187–1197.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. *21*, 327–335.

Sanders, S.L., Garbett, K.A., and Weil, P.A. (2002). Molecular characterization of *Saccharomyces cerevisiae* TFIID. Mol. Cell. Biol. *22*, 6000–6013.

Sawa, C., Nedea, E., Krogan, N., Wada, T., Handatt, H., Greenblatt, J., and Buratowski, S. (2004). Bromodomain factor 1 (Bdf1) is phosphorylated by protein kinase CK2. Mol. Cell. Biol. *24*, 4734–4742.

Saxton, W.O., and Baumeister, W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. J. Microsc. *127*, 127–138.

Segil, N., Guermah, M., Hoffmann, A., Roeder, R.G., and Heintz, N. (1996). Mitotic regulation of TFIID: inhibition of activator-dependent transcription and changes in subcellular localization. Genes Dev. *10*, 2389–2400.

Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.S., Buratowski, S., and Tan, S. (2001). A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. Nat. Struct. Biol. *8*, 695–700.

Stark, H., Mueller, F., Orlova, E.V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R., and van Heel, M. (1995). The 70S *Escherichia coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. Structure *3*, 815–821.

Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. Science *295*, 1058–1062.

Taatjes, D.J., Schneider-Poetsch, T., and Tjian, R. (2004). Distinct conformational states of nuclear receptor-bound CRSP-Med complexes. Nat. Struct. Mol. Biol. *11*, 664–671.

Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature *381*, 127–151.

Tanese, N., Pugh, B.F., and Tjian, R. (1991). Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. *5*, 2212–2224.

Tilley, S.J., Orlova, E.V., Gilbert, R.J., Andrew, P.W., and Saibil, H.R. (2005). Structural basis of pore formation by the bacterial toxin pneumolysin. Cell *121*, 247–256.

Tsai, F.T., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J. *19*, 25–36.

Unser, M., Trus, B.L., and Steven, A.C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. Ultramicroscopy *23*, 39–51.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Zhou, T., and Chiang, C.M. (2002). Sp1 and AP2 regulate but do not constitute TATA-less human TAF(II)55 core promoter activity. Nucleic Acids Res. *30*, 4145–4157.

Zhou, Z.H., Liao, W., Cheng, R.H., Lawson, J.E., McCarthy, D.B., Reed, L.J., and Stoops, J.K. (2001). Direct evidence for the size and conformational variability of the pyruvate dehydrogenase complex revealed by three-dimensional electron microscopy. The "breathing" core and its functional relationship to protein dynamics. J. Biol. Chem. *276*, 21704–21713.

Accession Numbers

The 3D reconstruction maps have been deposited at the EM Data Bank with the accession numbers 1194, 1195, and 1196.